Composante du produit nº 82-003-X au catalogue de Statistique Canada
Rapports sur la santé

Article

Évaluation de la couverture des données de l'Enquête sur la santé dans les collectivités canadiennes couplées aux dossiers de malades hospitalisés

par Michelle Rotermann

Janvier 2009

Statistique Canada Statistics Canada

Canada

Évaluation de la couverture des données de l'Enquête sur la santé dans les collectivités canadiennes couplées aux dossiers de malades hospitalisés

par Michelle Rotermann

Résumé

Contexte

L'évaluation de la couverture réalisée grâce au couplage de données administratives recueillies régulièrement par les hôpitaux et de données d'enquête est une importante activité préparatoire aux analyses fondées sur le fichier couplé.

Données et méthodes

Afin d'évaluer la couverture du couplage des données du cycle 1.1 de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) et de données sur des malades hospitalisés (Information-santé orientée vers la personne, ou ISOP), le nombre de personnes admises à l'hôpital selon l'ISOP a été comparé à l'estimation pondérée des répondants de l'ESCC qui ont pu être couplés aux données de l'ISOP. Les écarts entre l'ISOP et l'estimation couplée et pondérée de l'ESCC a révélé l'échec du couplage ou une couverture insuffisante.

Résultats

Selon l'ISOP, 1 572 343 personnes (hors du Québec) âgées de 12 ans et plus ont été hospitalisées de septembre 2000 à la fin novembre 2001. Les estimations pondérées tirées de l'ESCC, rajustées pour tenir compte du consentement au couplage et de numéros d'assurance-maladie vraisemblables, étaient inférieures de 7,7 % aux données de l'ISOP. Les taux de couverture des hommes et des femmes étaient comparables. De plus, les taux provinciaux ne différaient pas de ceux des autres régions du Canada, quoique les territoires aient affiché des différences. Les taux de couverture étaient sensiblement plus faibles dans la tranche des 75 ans et plus qu'ils ne l'étaient parmi les personnes âgées de 12 à 74 ans.

Mots-clés

Couverture, collecte de données, bases de données, enquêtes sur la santé, dossiers d'hôpital, malades hospitalisés, dossiers médicaux, couplage probabiliste.

Auteure

Michelle Rotermann (1-613-951-3166; Michelle.Roterman@statcan.gc.ca) travaille à la Division de l'information et de la recherche sur la santé, à Statistique Canada, Ottawa (Ontario) K1A 0T6.

Le couplage d'enregistrements est utilisé dans les études sur la santé pour recueillir de l'information complète, combler des écarts dans des ensembles de données actuelles ou améliorer la qualité des données[1,2]. Notamment, grâce à la confirmation du décès de participants à des enquêtes, aux cohortes visées par des études, ou aux sources de données administratives, tels les dossiers de malades hospitalisés, il a été possible d'étudier les rapports entre le décès et des facteurs comme le mode de vie, la profession, les modalités de traitement, les antécédents médicaux et la géographie[3-20].

Dans la même optique, le couplage d'enregistrements de naissances et de mortinaissances et d'enregistrements de décès ou d'hospitalisations a permis d'étudier la morbidité et la mortalité des mères, des fœtus et des nourrissons selon les caractéristiques de la mère et du nourrisson[21-24]. La technique sert également à valider de l'information autodéclarée[25,26], à décrire les caractéristiques d'enregistrements non appariés[27], à évaluer la comparabilité ou la qualité de fichiers de données générées par des méthodes de couplage probabiliste et déterministe[28], à réduire la sous-détermination de la prévalence de maladies[29] et à contrôler le rendement du système de santé[30,31]. En l'absence de registres de maladies, le couplage d'enregistrements constitue un moyen économique et efficient de surveiller l'incidence et la prévalence de celles-ci[32-35].

La présente étude a été motivée par la nécessité d'évaluer la couverture du couplage de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) et de l'Information-santé orientée vers la personne (ISOP), *une base de données administrative de*

dossiers d'hôpital. La recherche initiale sur le taux de couplage de l'ESCC et de l'ISOP a consisté à estimer la part des participants à l'ESCC qui avaient été hospitalisés entre 1994-1995 et 2004-2005, mais il reste à évaluer la couverture du fichier couplé[36]. L'évaluation de la couverture est indispensable si le fichier couplé doit servir à la recherche épidémiologique. Il importe de déterminer si les conclusions seront biaisées, à savoir si les participants qui possèdent certaines caractéristiques sont plus susceptibles d'être couplés que ceux qui ne les possèdent pas.

L'ISOP et l'ESCC sont des sources de données complémentaires. L'ISOP ne renferme pas de renseignements sur les déterminants non médicaux de la santé, comme les facteurs socioéconomiques et les facteurs liés au mode de vie. Ainsi, les dossiers d'hôpital ne renseignent aucunement sur l'usage du tabac ou l'indice de masse corporelle (IMC), deux facteurs de risque importants. En revanche, l'ESCC est riche en information sur l'état de santé et les déterminants de la santé, mais pauvre en détails nécessaires à l'étude de l'hospitalisation. La fonte de l'ISOP et de l'ESCC réduit bon nombre des limites de chaque source et facilite dès lors une compréhension approfondie des raisons pour lesquelles les Canadiens sont mis en contact avec le système de soins de santé et de leurs résultats lorsqu'ils se trouvent dans le système.

Les deux principaux objectifs de la présente étude étaient : 1) de calculer le taux de couverture de l'ESCC et de l'ISOP couplées; 2) de cerner les caractéristiques des participants au cycle 1.1 de l'ESCC dont il était moins probable que les autres qu'ils se trouvent dans le fichier couplé.

Méthodes

Sources de données

Enquête sur la santé dans les collectivités canadiennes

L'Enquête sur la santé dans les collectivités canadiennes (ESCC) est une enquête transversale qui recueille de l'information sur l'état de santé, le recours aux soins de santé et les déterminants de la santé. Elle a pour champ d'observation la population à domicile de 12 ans et plus des provinces et des territoires, sauf les membres réguliers des Forces canadiennes, les personnes vivant en établissement et les résidents des réserves indiennes, des autres établissements autochtones et de certaines régions éloignées. Son taux de couverture a été de 98 % dans les provinces, de 97 % dans les Territoires du Nord-Ouest, de 90 % au Yukon et de 71 % au Nunavut.

Les données du cycle 1.1 ont été recueillies du 1er septembre 2000 au 3 novembre 2001 inclusivement auprès d'un échantillon de 131 535 personnes. Le taux de réponse a été de 84,7 %. Tous les renseignements donnés en réponse à l'ESCC, notamment les numéros des régimes provinciaux d'assurance-maladie et les codes postaux, sont autodéclarés par les répondants, et l'ampleur de l'erreur de ces variables est inconnue. Cependant, les applications de saisie qu'utilisent les intervieweurs comprennent des fonctions de dépistage des réponses incohérentes ou non incluses dans la fourchette de valeurs permises et des séquences alphanumériques non valables. Un rapport publié offre de plus amples renseignements sur l'ESCC[37].

Les participants à l'ESCC ont été invités à autoriser le couplage de l'information recueillie lors de l'interview et d'information provinciale sur leur santé, notamment le recours antérieur et actuel à des services tels ceux d'hôpitaux, de cliniques ou de cabinets de médecin ou à d'autres services fournis par la province; 91 % ont accordé l'autorisation demandée. L'échantillon utilisé aux fins de la présente étude réunit 72 354 (66,5 %) personnes âgées de 12 ans et plus dans toutes les provinces et tous les territoires sauf le Québec qui ont consenti au couplage et fourni un numéro d'assurance-maladie (NAM) valide (annexe A). Les enregistrements de l'ISOP concernant le Québec ne peuvent pas être couplés à ceux de l'ESCC, car les dossiers d'hôpital du Québec fournis à Statistique Canada contiennent des numéros d'assurance-maladie brouillés, des codes postaux incomplets et ne précisent aucune date de naissance.

Des poids d'enquête ont été utilisés afin que les estimations tirées des données de l'ESCC soient représentatives de la population cible plutôt que du seul échantillon. « Poids d'enquête » s'entend du nombre de personnes contenues dans la population que représente chaque répondant. Il témoigne des différentes probabilités de sélection et de réponse. Par conséquent, chaque enregistrement est pondéré de l'inverse de la probabilité de sélection de la personne et de recueil de sa réponse[38]. Le couplage d'enregistrements nécessite le recours à des poids d'enquête supplémentaires, car les répondants ne consentent pas tous au couplage, et ceux qui le font ne fournissent pas tous un NAM valide. Aux fins de la présente étude, des poids d'enquête ajustés pour tenir compte du consentement au couplage et de la déclaration d'un NAM valide ont été calculés.

Statistique Canada n'a pas accès aux bases de données de l'assurance-maladie provinciale dans lesquelles vérifier les NAM déclarés par les répondants de l'ESCC. À la place, les provinces et les territoires sans exception ont fourni des formules de contrôle numérique qui confirment que les NAM sont au moins vraisemblables. Si les chiffres de contrôle ne peuvent pas remplacer des bases de données qui contiennent noms et prénoms, dates de naissance, adresses et NAM, ils permettent quand même de déceler les erreurs de

transcription, par exemple l'inversion de deux chiffres, et constituent un moyen simple de distinguer les numéros valables de séries de chiffres aléatoires.

Données d'hôpital

La base de données de l'Information-santé orientée vers la personne (ISOP) tenue par Statistique Canada contient de l'information sur la radiation des patients (pour cause de congé et de décès à l'hôpital) de la quasi-totalité des hôpitaux de soins actifs et de quelques hôpitaux psychiatriques, hôpitaux pour malades chroniques et établissements de réadaptation.

L'ISOP est un ensemble de données au niveau de la personne constitué des enregistrements des congés (parfois des congés multiples donnés à la même personne) contenus dans la Base de données sur la morbidité hospitalière (BDMH). Les antécédents d'hospitalisation de chaque malade peuvent être établis d'après les enregistrements séquentiels au niveau de la personne de l'ISOP. Au cours du couplage, les enregistrements se rapportant à un même particulier sont repérés au moyen du NAM, d'information démographique et de renseignements sur les diagnostics et les interventions (sexe, date de naissance, procédures propres à un sexe, et autres)[39].

Les dossiers d'hôpital de l'exercice écoulé sont ajoutés chaque année à l'ISOP. Chaque fois que les données d'une année sont entrées dans le système, le processus complet de l'ISOP est exécuté afin de garantir la cohérence interne de l'information démographique au niveau de la personne concernant les patients ayant reçu de multiples congés.

Des études de données hospitalières obtenues par une seconde saisie, qui visent à confirmer l'exactitude de celles-ci, ont démontré la grande qualité des éléments de données administratifs et non médicaux (essentiels au couplage d'enregistrements). Ainsi, 99 % d'un échantillon aléatoire d'enregistrements de congé à la suite d'un séjour à l'hôpital de septembre à la fin novembre 2000 contenait des NAM exacts, et 91 % des codes postaux étaient sans erreur[40].

Statistique Canada possède des données d'hôpital comprenant les NAM pour toutes les provinces (sauf le Québec) et les Territoires du Nord-Ouest de 1994-1995 à ce jour, et des données de même type pour certaines provinces de 1992-1993 et 1993-1994.

Bien que la base de données de l'ISOP comprenne la vaste majorité des enregistrements de la BDMH, quelque 3 % de ceux des malades âgés de 12 ans et plus (la population cible de la présente étude) en étaient exclus à cause de NAM manquants ou non valides[39].

Du 1er septembre 2000 au 3 novembre 2001 inclusivement, 1 624 972 personnes âgées de 12 ans et plus ont reçu 2,3 millions de congés d'hôpitaux de soins actifs hors du Québec. L'étude ne s'étendait pas aux congés d'hôpitaux de soins autres que de soins actifs, car leur couverture n'est pas uniforme d'une province à l'autre.

Les populations cibles de l'ESCC et de l'ISOP diffèrent quelque peu. Par exemple, l'ESCC exclut les membres à plein temps des Forces canadiennes de même que les personnes vivant dans les réserves indiennes, en établissement (p. ex. foyers de soins infirmiers et prisons) et dans certaines régions éloignées. L'ISOP est un recensement

Figure 1
Évaluation et pointage types de paires d'enregistrements de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) et de l'Information-santé orientée vers la personne (ISOP) au moyen du Système généralisé de couplage d'enregistrements (SGCE)

ID	Province	Date de naissance	Code postal	Numéro d'assurance maladie (NAM)	Sexe	Pointage calculé par le SGE	Apparie-ment?	Commentaires
A	Ontario	1964-06-11	L9Y3B9	3512345678	Femmes	447	oui	Tous les champs correspondent.
1	Ontario	1964-06-11	L9Y3B9	3512345678	Femmes			
F	Manitoba	1927-07-24	R0A0T0	55667788	Hommes	-308	non	Aucun champ ne correspond.
1	Ontario	1964-06-11	L9Y3B9	3512345678	Femmes			
B	Manitoba	1945-05-21	R0A0T0	manquant	Hommes	268	oui	NAM manquant; tous les autres champs correspondent.
7	Manitoba	1945-05-21	R0A0T0	4624252627	Hommes			
B	Manitoba	1945-05-21	R0A0T0	manquant	Hommes	-244	non	NAM manquant; aucun autre champ ne correspond.
1	Ontario	1964-06-11	L9Y3B9	3512345678	Femmes			

☐ Enregistrement de l'ESCC ☐ Enregistrement de l'ISOP

qui dénombre les hospitalisations, y compris parmi ces populations. Par souci d'assurer la correspondance des populations de l'ESCC et de l'ISOP, les hospitalisations parmi les personnes vivant dans les réserves ou en établissement ont été exclues de l'analyse.

La population des réserves est une variable de recensement dérivée, constituée par classement des subdivisions de recensement (SDR) selon des critères établis par Affaires indiennes et du Nord Canada (AINC), et comprenant des SDR choisies parmi les collectivités septentrionales de la Saskatchewan, des Territoires du Nord-Ouest et du Yukon[41]. Les personnes vivant en établissement ont été identifiées par consultation du Fichier de conversion des codes postaux (FCCP+)[42] et d'une liste d'établissements servant à l'Enquête sur les établissements de soins pour bénéficiaires internes[43]. Les hospitalisations de 31 330 personnes vivant dans les réserves et dans des régions apparentées ont été éliminées de l'ISOP, tout comme celles de 21 299 personnes vivant en établissement. L'élimination des 52 629 enregistrements en question, soit quelque 3 % des malades hospitalisés selon l'ISOP au cours de la période à l'étude, a eu pour résultats d'adapter la population visée par l'ISOP à la population cible de l'ESCC.

Tableau 1
Nombre de personnes hospitalisées dans des hôpitaux de soins actifs et taux de couverture, du 1er septembre 2000 au 3 novembre 2001 inclusivement, selon des caractéristiques choisies et la source de données, population des personnes âgées de 12 ans et plus, le Canada excluant le Québec

	Information-santé orientée vers la personne (ISOP)	Enquête sur la santé dans les collectivités canadiennes (ESCC)		Taux de couverture ESCC-ISOP		
					Intervalle de confiance de 95 %	
	Nombre	Nombre non pondéré	Nombre pondéré	%	de	à
Total	**1 572 343**	**6 785**	**1 451 272**	**92,3**	**88,9**	**95,7**
Province/Territoires						
Terre-Neuve-et-Labrador	41 394	272	40 445	97,7	83,6	111,8
Île-du-Prince-Édouard	11 784	237	11 061	93,9	79,6	108,1
Nouvelle-Écosse	67 226	348	60 419	89,9	78,0	101,7
Nouveau-Brunswick	67 542	423	62 203	92,1	81,7	102,5
Ontario	753 970	2 230	694 463	92,1	86,6	97,6
Manitoba	82 386	567	69 739	84,6	73,6	95,7
Saskatchewan	82 778	659	78 664	95,0	86,4	103,7
Alberta	202 498	863	186 301	92,0	83,3	100,7
Colombie-Britannique	258 883	1 062	241 647	93,3	85,3	101,3
Territoires	3 882	124	6 331	163,1*	139,3	186,9
Sexe						
Femmes†	971 087	4 343	904 318	93,1	88,8	97,5
Hommes	601 249	2 442	546 955	91,0	85,4	96,5
Groupe d'âge						
12 à 24 ans†	173 009	711	165 093	95,4	81,5	101,3
25 à 34 ans	276 150	1 041	270 164	97,8	85,0	105,8
35 à 44 ans	210 848	830	209 447	99,3	88,3	110,4
45 à 54 ans	183 878	773	174 194	94,7	84,9	104,6
55 à 64 ans	181 041	854	169 235	93,5	83,2	103,7
65 à 74 ans	227 410	1 090	219 259	96,4	86,9	106,0
12 à 74 ans	1 252 336	5 299	1 207 392	96,4	92,4	100,4
75 ans et plus	320 007	1 486	243 881	76,2*	70,2	82,2

† catégorie de référence
* valeur significativement différente de celle pour la catégorie de référence ($p < 0,05$), aux fins de la comparaison provinciale, valeur significativement différente de celle pour le reste du Canada, par exemple, l'Ontario comparé au Canada moins l'Ontario
Source : Enquête sur la santé dans les collectivités canadiennes, 2000-2001, Information-santé orientée vers la personne, de 2000-2001 à 2001-2002

Techniques d'analyse

Couplage d'enregistrements probabiliste

Les participants à l'ESCC qui avaient été hospitalisés ont été identifiés par couplage d'enregistrements probabiliste. Le couplage de l'ESCC et de l'ISOP a été effectué au moyen du Système généralisé de couplage d'enregistrements (SGCE), un logiciel développé par Statistique Canada. Si les deux sources de données contiennent de nombreuses variables, elles n'ont que quelques champs en commun, lesquels sont suffisamment distincts pour effectuer l'appariement. Un participant à l'ESCC n'était réputé avoir été hospitalisé que si l'ISOP contenait un enregistrement comprenant un NAM ou des caractéristiques démographiques comparables (date de naissance, sexe, code postal) et précisait une date d'admission à un établissement de soins actifs entre le 1er septembre 2000 et le 3 novembre 2001.

Le couplage probabiliste ne nécessite pas la correspondance complète des variables d'appariement. La qualité de l'appariement des enregistrements est plutôt évaluée par l'application d'algorithmes qui mesurent la probabilité d'un appariement correct[1-44] (figure 1). Des points étaient attribués ou déduits, selon la similitude des valeurs des champs. À titre d'exemple, un pointage positif élevé était attribué là où les NAM et les provinces de délivrance étaient identiques; si les valeurs étaient comparables, sans toutefois être exactes, un pointage positif moindre était accordé pour tenir compte de la correspondance partielle; des points étaient déduits si les valeurs contenues dans les deux enregistrements divergeaient carrément.

Le nombre de points alloués à chaque paire de variables de couplage était à proportion de leur importance en tant que variables d'appariement, ce qui était normalement fonction de leur

unicité. Ainsi, comme il n'existe que deux valeurs possibles du sexe du participant/malade, le nombre de points attribués à la correspondance de ces valeurs était inférieur au pointage accordé si les codes postaux ou les NAM correspondaient.

Le poids de couplage total attribué à chaque paire d'enregistrements ESCC-ISOP était calculé en additionnant les points alloués à chaque paire de variables d'appariement. Plus le poids de couplage total était élevé, plus il était probable que les deux enregistrements concernaient la même personne. Les poids de couplage totaux répondent idéalement à une distribution bimodale. Les paires d'enregistrements qui recueillaient un pointage supérieur au seuil choisi étaient considérées comme des appariements vrais, tandis que celles dont le poids y était inférieur étaient rejetées. Pour éviter de devoir effectuer un examen manuel, les analystes ont fixé des seuils d'exclusion identiques aux fins de l'étude, si bien que l'une seulement des deux valeurs (appariement et non-appariement) pouvait être attribuée à chaque paire d'enregistrements.

Résultats

Afin d'évaluer la couverture du couplage du cycle 1.1 de l'ESCC et de l'ISOP, le nombre de personnes admises à l'hôpital selon chaque source de données a été comparé. Des poids de recensement, ajustés en fonction du consentement au couplage et de la validité du NAM, ont été appliqués aux enregistrements des participants à l'ESCC pour lesquels des données ont été repérées dans la base de données de l'ISOP. Le nombre d'hospitalisations selon l'ISOP était considéré comme la norme. Le taux de couverture a été calculé en divisant le nombre estimatif pondéré de participants à l'ESCC dont des données ont été repérées dans l'ISOP par le nombre selon l'ISOP, moins les enregistrements des personnes vivant dans les réserves indiennes ou des régions apparentées ou en établissement, puis en multipliant le quotient par 100. Les écarts entre les données de l'ISOP et les estimations pondérées de l'ESCC ont été examinés. Des erreurs-types et des intervalles de confiance de 95 % ont été calculés pour les taux de couverture par la méthode du *bootstrap*. La signification statistique a été soumise à un test t ($p < 0,05$)[45,46].

Selon l'ISOP, 1 572 343 personnes ont été admises à un hôpital de soins actifs (excluant ceux du Québec), du 1er septembre 2000 au 3 novembre 2001 inclusivement (tableau 1). Les estimations pondérées de l'ESCC, rajustées pour tenir compte du consentement au couplage et d'un NAM valide, étaient inférieures de 7,7 % (1 451 272) à cette donnée.

Les taux de couverture étaient comparables pour les hommes et pour les femmes (91,0 % et 93,1 %). Les taux provinciaux ne différaient pas significativement du taux pour le reste du Canada. Toutefois, le nombre estimatif d'habitants des territoires qui ont été hospitalisés selon l'ESCC était significativement plus élevé que la donnée de l'ISOP. Par conséquent, le taux de couverture des territoires est supérieur à 100 %.

Les taux de couverture de la plupart des groupes d'âge étaient comparables, à ceci près que celui des 75 ans et plus (76,2 %) était passablement inférieur à la donnée correspondante pour la tranche des 12 à 74 ans (96,4 %).

Discussion

Le taux de couverture significativement moindre des personnes âgées de 75 ans et plus était prévu, car les deux sources de données ne concernent pas des populations identiques. L'ESCC exclut les personnes vivant en établissement, lesquelles sont incluses dans les données sur l'hospitalisation (ISOP). Or, les résidents d'établissements sont beaucoup plus nombreux parmi les personnes âgées que parmi les jeunes : dans l'ensemble, moins de 2 % des Canadiens entrent dans cette catégorie, donnée qui passe à 16 % pour la tranche des 75 ans et plus[47].

À défaut de renseignements directs dans les enregistrements de l'ISOP sur le lieu de résidence des malades, le code postal conjugué au FCCP+ et à la liste des établissements de soins pour bénéficiaires internes a servi à déterminer si les malades vivaient en établissement. De cette façon, plus de 20 000 résidents d'établissements ont été identifiés et supprimés de l'ISOP au moyen du FCCP+. Néanmoins, le taux de couverture des personnes âgées de 75 ans et plus est demeuré significativement inférieur aux taux se rapportant aux personnes plus jeunes.

Le recours au FCCP+ et à la liste des établissements de soins pour bénéficiaires internes conjugués au seul code postal n'est pas idéal. Les établissements habités par la majorité de la population dont le code postal était le même étaient plus susceptibles que les autres d'être identifiés, puis éliminés des données de l'ISOP. Qui plus est, les codes postaux des établissements en milieu urbain sont distinctifs, et la probabilité que leurs résidents soient éliminés de l'ISOP était supérieure à la normale. Bien souvent, le même code postal est attribué à de nombreux secteurs de dénombrement et aires de diffusion situés en milieu rural, dans la banlieue éloignée ou dans de petites villes. Leur codage est donc beaucoup moins précis que ne l'est celui des zones urbaines centralisées, lesquelles sont normalement rattachées à un seul secteur de dénombrement ou à une seule aire de diffusion. Il s'ensuit que les données de l'ISOP ont probablement continué à tenir compte des résidents d'établissements de milieux ruraux, de la banlieue éloignée et de petites villes.

Le taux de couverture dans les territoires pose également problème, étant donné que les estimations couplées ESCC-ISOP ont dépassé la norme (ISOP). Toutefois, cette situation n'est pas inquiétante, car le petit nombre d'enregistrements de l'ESCC couplés à l'ISOP (124) écarte le risque d'analyses

futures englobant cette sous-population. Le taux de couverture des territoires était de 113 % avant que la population des réserves soit éliminée des données de l'ISOP, après quoi il est passé à 163 %. Il reste à déterminer pourquoi l'estimation après couplage de l'ESCC et de l'ISOP est de loin supérieure aux données de l'ISOP. Les enregistrements des participants à l'ESCC habitant les territoires ont été examinés afin de constater si des poids d'enquête élevés avaient été attribués à certains d'entre eux, ce qui pourrait expliquer l'écart entre les données de l'ISOP et celles issues du couplage ESCC-ISOP. Aucun poids divergent n'a été constaté : la moyenne était de 51, et les poids se situaient tous dans la fourchette de 11 à 178.

Par ailleurs, il était impossible de cerner les hospitalisations de personnel militaire afin de les éliminer de l'ISOP. Les membres à plein temps des Forces canadiennes sont exclus de l'ESCC, et leur prise en compte pourrait influer sur le taux de couverture.

Conclusion

La valeur du couplage d'enregistrements pour les études épidémiologiques de la santé de la population n'est plus à démontrer. L'appariement d'information tirée de données administratives sur la santé recueillies régulièrement, comme celles de l'ISOP, et de données d'enquête, comme les résultats de l'ESCC, nous procure des enseignements riches sur les déterminants de la santé, diverses catégories de recours aux soins de santé et les résultats en matière de santé. Or, l'évaluation de la couverture est une condition préalable aux analyses qui regroupent de l'information sur la santé de sources multiples en prenant appui sur le fichier couplé ESCC-ISOP.

La présente évaluation démontre que le taux de couverture global est élevé, dépassant souvent les 90 %, bien qu'il soit plus faible dans le cas de certains participants à l'ESCC, notamment les personnes âgées. Cependant, cette restriction n'est pas un obstacle, pourvu que les utilisateurs du fichier reconnaissent explicitement que les conclusions ne concernent que la population générale des ménages (population cible de l'ESCC) et non la population totale, particulièrement les personnes vivant en établissement.

Remerciements

L'auteure remercie Claude Nadeau de son aide et de ses judicieux commentaires.

Références

1. M.E. Fair et P. Whitridge, *Tutorial on Record Linkage* [diaporama], 1997 Federal Committee on Statistical Methodology, disponible à l'adresse http://www.fcsm.gov/working-papers/RLT97_chap12.html.
2. M. Fair, « Generalized record linkage system – Statistics Canada's record linkage software », *Austrian Journal of Statistics*, 33(1 et 2), 2004, p. 37-53.
3. P.J. Villeneuve, H.I. Morrison, C.L. Craig *et al.*, « Physical activity, physical fitness, and risk of dying », *Epidemiology*, 9(6), 1998, p. 626-631.
4. I. Weller et P. Corey, « The impact of excluding non-leisure energy expenditure on the relation between physical activity and mortality in women », *Epidemiology*, 9(6), 1998, p. 632-635.
5. I.R. Bata, R.D. Gregor, B.J. Eastwood *et al.*, « Tendances dans l'incidence de l'infarctus aigu du myocarde entre 1984 et 1993 : le projet MONICA du comté de Halifax », *Journal de la société canadienne de cardiologie / Canadian Journal of Cardiology*, 16(5), 2000, p. 589-595.
6. N.P. Roos et B. Havens, « Predictors of successful aging: a twelve-year study of Manitoba elderly », *American Journal of Public Health*, 81(1), 1991, p. 63-68.
7. K.T. Khaw, N. Wareham, S. Bingman *et al.*, « Combined impact of health behaviors and mortality in men and women: the EPIC-Norfolk prospective population study », *PLOS Medicine*, 2008, disponible à l'adresse http://medicine.plosjournals.org/perlserv/?request=get-document&doi10.1371/journal.pmed.0050012.
8. T.L. Weston, K.J. Aronson, G.R. Howe *et al.*, « Cancer mortality among males in relation to exposures assessed through a job-exposure matrix », *International Journal of Occupational and Environmental Health*, 6(3), 2000, p. 194-202.
9. K.J. Aronson, G.R. Howe, M. Carpenter *et al.*, « Surveillance of potential associations between occupations and causes of death in Canada, 1965-91 », *Occupational and Environmental Medicine*, 56(4), 1999, p. 265-269.
10. L. Wei, C.C. Lang, F.M. Sullivan *et al.*, « Impact on mortality following first acute myocardial infarction of distance between home and hospital: Cohort study », 2008, disponible à l'adresse http://www.heart.bmj.com/onlinefirst.dtl.
11. H.N. Schmaltz, D. Southern, W.A. Ghali *et al.*, « Living alone, patient sex and mortality after acute myocardial infarction », *Journal of General Internal Medicine*, 22(5), 2007, p. 572-578.

12. L.F. Ellison, L. Gibbons et le Groupe d'analyse de la survie au cancer au Canada, « Taux relatifs de survie à cinq ans : cancers de la prostate, du sein, du côlon et du rectum, et du poumon », *Rapports sur la santé*, 13(1), 2001, p. 25-34 (Statistique Canada, nº 82-003 au catalogue).

13. L.F. Ellison et L. Gibbons « Les principaux cancers – évolution de la survie relative à cinq ans », *Rapports sur la santé*, 15(2), 2004, p. 19-33 (Statistique Canada, nº 82-003 au catalogue).

14. P. Kaul, P.W. Armstrong, W. Change *et al.*, « Long-term mortality of patients with acute myocardial infarction in the United States and Canada: Comparison of patients enrolled in global utilization of streptokinase and t-PA for occluded coronary arteries (GUSTO)-I », *Circulation*, 110, 2004, p. 1754-1760, disponible à l'adresse http://circ.ahajournals.org/cgi/content/full/110/13/1754.

15. C.M. Fischbacher, R. Bhopal, C. Povey *et al.*, « Record linked retrospective cohort study of 4.6 million people exploring ethnic variations in disease: myocardial infarction in South Asians », *BMC Public Health*, 7, 2007, p. 142.

16. L.S. Jebamani, C.A. Burchill et P.J. Martens, « Using data linkage to identify First Nations Manitobans: technical, ethical, and political issues », *Canadian Journal of Public Health / Revue canadienne de santé publique*, 96(Suppl.), 2005, p. S28-S32.

17. K.T. Knoops, L.C. de Groot, D. Kromhout *et al.*, « Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project », *The Journal of the American Medical Association*, 292, 2004, p. 1433-1439.

18. P.K. Myint, A.A. Welch, S.A. Bingham *et al.*, « Fruit and vegetable consumption and self-reported functional health in men and women in the European Prospective Investigation into Cancer-Norfolk (EPIC-Norfolk): a population-based cross-sectional study », *Public Health Nutrition*, 10(1), 2007, p. 34-41.

19. A. Robert Schnatter, « An analysis of death ascertainment and follow-up through Statistics Canada's Mortality Data Base System », *Canadian Journal of Public Health / Revue canadienne de santé publique*, 81(1), 1990, p. 60-65.

20. J.A. Schoenman, J.P. Sutton, A. Elixhauser *et al.*, « Understanding and enhancing the value of hospital discharge data », *Medical Care Research and Review*, 64(4), 2007, p. 449-468.

21. B. Herrchen, J.B. Gould et T.S. Nesbitt, « Vital statistics linked birth/infant death and hospital discharge record linkage for epidemiological studies », *Computing Biomedical Research*, 30(4), 1997, p. 290-305.

22. J. Chen, M. Fair, R. Wilkins *et al.*, « Niveau de scolarité de la mère et mortalité fœtale et infantile au Québec », *Rapports sur la santé*, 10(2), 1998, p. 57-71 (Statistique Canada, nº 82-003 au catalogue).

23. S.W. Wen, M.S. Kramer, S. Liu *et al.*, « Mortalité infantile d'après l'âge gestationnel et le poids à la naissance dans les provinces et territoires au Canada, naissances entre 1990 et 1994 », *Maladies chroniques au Canada / Chronic Diseases in Canada*, 21(1), 2000, p. 14-22.

24. M.S. Kramer, K. Demissie, H. Yang *et al.*, « The contribution of mild and moderate preterm birth to infant mortality », *Journal of the American Medical Association*, 284(7), 2000, p. 843-849.

25. S.A. Bingham, M. Cassidy, T.J. Cole *et al.*, « Validation of weighed records and other methods of dietary assessment using the 24h urine nitrogen technique and other biological markers », *British Journal of Nutrition*, 73, 1995, p. 531-550.

26. M.A. Quigley, C. Hockley et L.L. Davidson, « Agreement between hospital records and maternal recall of mode of delivery: evidence from 12 391 deliveries in the UK Millennium Cohort Study », *International Journal of Obstetrics and Gynaecology*, 114(2), 2007, p. 195-200.

27. J.B. Ford, B. Herrchen, J.B. Gould *et al.*, « Characteristics of unmatched maternal and baby records in linked birth and hospital discharge data », *Health Statistics Quarterly*, 33, 2007, p. 25-33.

28. M. Fair, M. Cyr, A. Allen *et al.*, *Étude de validation d'un couplage d'enregistrements de naissance et de décès infantile au Canada*, 1999 (Statistique Canada, nº 84-F0013 au catalogue).

29. D.G. Manuel et S.E. Schultz, « Using linked data to calculate summary measures of population health: Health-adjusted life expectancy of people with Diabetes Mellitus », *Population Health Metrics*, 2(4), 2004, disponible à l'adresse http://www.pophealthmetrics.com/content/2/1/4 (document consulté le 13 mai 2008).

30. J. Calver, K.J. Brameld, D.B. Preen *et al.*, « High-cost users of hospital beds in Western Australia: a population-based record linkage study », *Medical Journal of Australia*, 184(8), 2006, p. 393-397.

31. C. Houle, J.M. Berthelot, P. David *et al.*, *Le projet d'appariement du Recensement et des fichiers de soins de santé du Manitoba*, document de recherche nº 91, Direction des études analytiques, 1996, disponible à l'adresse http://www.fcsm.gov/working-papers/choule.pdf (Statistique Canada, nº 11F0019MPF au catalogue).

32. S. Brocco, C. Visentin, U. Fedeli *et al.*, « Monitoring the occurrence of diabetes mellitus and its major complications: the combined use of different administrative databases », *Cardiovascular Diabetology*, 6(5), 2007.

33. T.K. Young, E. Kliewer, J. Blanchard *et al.*, « Monitoring disease burden and preventive behavior with data linkage: cervical cancer among aboriginal people in Manitoba, Canada », *American Journal of Public Health*, 90(9), 2000, p. 1466-1468.

34. C. Ringland, P.K. Correll, K. Lim *et al.*, « Hospital readmissions for asthma: a feasibility study comparing strategies for linking hospital morbidity data », *Australian and New Zealand Journal of Public Health*, 30(5), 2006, p. 435-439.

35. K.J. Aronson, G.R. Howe, M. Carpenter *et al.*, *Dépistage des maladies et lésions professionelles au Canada : mortalité par cause chez les travailleurs, 1965-1991* [CD-ROM], 2000 (Statistique Canada, nº 84-546 au catalogue).

36. C. Nadeau, M.P. Beaudet et J. Marion, « Couplage d'enregistrements déterministe et probabiliste », *Symposium 2006 de Statistique Canada : Enjeux méthodologiques reliés à la mesure de la santé des populations*, disponible à l'adresse www.statcan.ca/francais/freepub/11-522-XIF/2006001/article/10404-fr.pdf.

37. Y. Béland, « Enquête sur la santé dans les collectivités canadiennes – aperçu de la méthodologie », *Rapports sur la santé*, 13(3), 2002, p. 9-15 (Statistique Canada, n° 82-003 au catalogue).

38. Statistique Canada, Enquête sur la santé dans les collectivités canadiennes, Sources de données et méthodologie, Estimation, disponible à l'adresse www.statcan.ca/cgi-bin/imdb/p2SV_f.pl?Function=getSurvey&SDDS=3226&lang=en&db=IMDB&dbg=f&adm=8&dis=2#b6 (document consulté le 27 juin 2008).

39. Statistique Canada, *External linkage production report: Data years: F1992 to F2004* [rapport], Division des méthodes d'enquête auprès des ménages, 2006 (non publié).

40. J. Richards, A. Brown et C. Homan, « L'étude de la qualité des données de la base canadienne de données sur les congés des patients », *Symposium 2001 de Statistique Canada : La qualité des données d'un organisme statistique : une perspective méthodologique*, disponible à l'adresse www.statcan.ca/francais/freepub/11-522-XIF/2001001/session16/s16a_f.pdf (document consulté le 2 juillet 2008).

41. Statistique Canada, *Dictionnaire du recensement de 2001* (Statistique Canada, n° 92-378 au catalogue), disponible à l'adresse www12.statcan.ca/francais/census01/Products/Reference/dict/appendices/92-378-XIF02002.pdf (document consulté le 26 juin 2008).

42. Statistique Canada, *Fichier de conversion des codes postaux (FCCP): mise à jour*, disponible à l'adresse www.statcan.ca/bsolc/francais/bsolc?catno=92-153-UCB (document consulté le 28 juin 2008).

43. Statistique Canada, *Établissements de soins pour bénéficiaires internes (ESBI)*, disponible à l'adresse www.statcan.ca/bsolc/francais/bsolc?catno=83-237-X (document consulté le 2 juillet 2008).

44. I.P. Fellegi et A.B. Sunter, « A theory for record linkage », *Journal of the American Statistical Association*, 64, 1969, p. 1183-1210.

45. J.N.K. Rao, C.F.J. Wu et K. Yue, « Quelques travaux récents sur les méthodes de rééchantillonnage applicables aux enquêtes complexes », *Techniques d'enquête*, 18(2), 1992, p. 225-234 (Statistique Canada, n° 12-001 au catalogue).

46. K.F. Rust et J.N.K. Rao, « Variance estimation for complex surveys using replication techniques », *Statistical Methods in Medical Research*, 5(3), 1996, p. 281-310.

47. M. Turcotte et G. Schellenberg, *Un portrait des aînés au Canada*, Ottawa, Ministre de l'industrie, 2007, disponible à l'adresse www.statcan.ca/francais/freepub/89-519-XIF/89-519-XIF2006001.pdf (document consulté le 4 janvier 2008) (Statistique Canada, n° 89-519 au catalogue).

Annexe

Annexe – Tableau A
Participants (nombre et pourcentage) à l'Enquête sur la santé dans les collectivités canadiennes qui ont consenti au couplage de leurs réponses à l'ESCC et de leur dossier de santé administratif et qui ont fourni un NAM valide, selon des caractéristiques choisies, le Canada excluant le Québec, 2001

	Consentement		Consentement et NAM valide	
	Nombre	%	Nombre	%
Total	**98 450**	**90,4**	**72 354**	**66,5**
Province/Territoires				
Terre-Neuve-et-Labrador	3 533	91,3	2 933	75,8
Île-du-Prince-Édouard	3 238	88,7	2 236	61,2
Nouvelle-Écosse	4 938	92,8	4 108	77,2
Nouveau-Brunswick	4 634	92,8	3 746	75,0
Ontario	35 674	90,8	24 917	63,4
Manitoba	7 653	90,4	5 552	65,5
Saskatchewan	7 417	92,6	6 142	76,7
Alberta	12 757	88,2	9 155	63,3
Colombie-Britannique	16 493	90,1	11 990	65,5
Territoires	2 113	83,9	1 575	62,6
Sexe				
Femmes†	52 865	90,5	40 334	69,1
Hommes	45 585	90,3	32 020	63,4
Groupe d'âge				
12 à 24 ans†	19 246	91,8	13 538	64,5
25 à 34 ans	14 482	90,9	10 119	63,5
35 à 44 ans	18 892	90,2	13 883	66,3
45 à 54 ans	16 036	89,6	11 905	66,5
55 à 64 ans	11 493	90,0	8 695	68,1
65 à 74 ans	9 778	90,3	7 684	71,0
12 à 74 ans	89 927	90,5	65 824	66,3
75 ans et plus	8 523	89,5	6 530	68,6

Source : Enquête sur la santé dans les collectivités canadiennes, 2000-2001